A PROXIMAL APPROACH FOR SIGNAL RECOVERY BASED ON INFORMATION MEASURES

Mireille El Gheche[1], *Anna Jezierska*[1], *Jean-Christophe Pesquet*[1] *and Joumana Farah*[2]

[1] Université Paris-Est
LIGM and UMR-CNRS 8049,
77454 Marne-la-Vallé, France
{mireille.elgheche, jeziersk, pesquet}@univ-mlv.fr

[2] Holy-Spirit University of Kaslik,
Faculty of Engineering, Telecom. Dep.,
P.O. Box 446, Jounieh, Lebanon
joumanafarah@usek.edu.lb

ABSTRACT

Recently, methods based on Non-Local Total Variation (NLTV) minimization have become popular in image processing. They play a prominent role in a variety of applications such as denoising, compressive sensing, and inverse problems in general. In this work, we extend the NLTV framework by using some information divergences to build new sparsity measures for signal recovery. This leads to a general convex formulation of optimization problems involving information divergences. We address these problems by means of fast parallel proximal algorithms. In denoising and deconvolution examples, our approach is compared with ℓ_2-NLTV based approaches. The proposed approach applies to a variety of other inverse problems.

Index Terms— Divergences, inverse problems, non-local processing, total variation, convex optimization, proximity operator, parallel algorithms.

1. INTRODUCTION

The motivation of this work is to investigate the influence of the choice of the smoothness measure on the performance of optimization approaches in the context of image recovery. More generally, we will consider a general class of convex optimization problems involving discrete information divergences. The purpose of these divergence terms often consists of enforcing prior knowledge or assumptions about the target solution. In this context, we are interested in an objective function having the following form:

$$x \mapsto D(Ax, Bx) \tag{1}$$

where D is a function in $\Gamma_0(\mathbb{R}^P \times \mathbb{R}^P)$, and A and B are matrices in $\mathbb{R}^{P\times N}$. Here $\Gamma_0(\mathcal{H})$ denotes the class of convex functions defined on a real Hilbert space $\mathcal{H}$, taking their values in $]-\infty, +\infty]$, and that are lower-semicontinuous and proper. The most common assumption on the signal of interest $\overline{x} \in \mathbb{R}^N$ is that, by making an appropriate choice of matrices A and B, $A\overline{x}$ and $B\overline{x}$ are close in the metric induced by D.

Non-Local Total Variation (NLTV) [1] has been used as a popular and effective image prior model in regularization-based imaging problems [2, 3, 4, 5]. It is known to reduce undesired staircase effects often present in Total Variation (TV) results. Indeed, the choice of directions pointed by local gradients is regarded as a drawback of TV prior. To circumvent this limitation, NLTV is associated with image-driven directions, i.e. the directions are chosen for each pixel independently, based on a similarity score between pixel intensities in a local neighborhood. In the following, we will consider generalizations of classical ℓ_2-NLTV. The ℓ_2-NLTV regularization term is a special case of (1), expressed as

$$D(Ax, Bx) = \sum_{n=1}^{N} \|A_n x - B_n x\| \tag{2}$$

where $\|\cdot\|$ is the ℓ_2-norm and $A = \left[A_1^\top, \ldots, A_N^\top\right]^\top$, $B = \left[B_1^\top, \ldots, B_N^\top\right]^\top$. More specifically, for every $n \in \{1, \ldots, N\}$, submatrix $A_n \in \mathbb{R}^{P_n\times N}$ has only one nonzero column, namely its n-th column is given by a vector of weights $(\omega^{(i)})_{1\le i\le P_n} \in [0, +\infty[^{P_n}$. Submatrix B_n has only one nonzero element in each row, i.e. the nonzero element in the i-th row is given by $\omega^{(i)}$. The couple of operators (A_n, B_n) is chosen adaptively for a given image in a preprocessing step. The nonzero elements of operator B_n correspond to components of x close to its n-th component $x^{(n)}$ in terms of some similarity measure. Usually, a patch based score [6] is considered. In the following, we examine functions different from ℓ_2 for defining NLTV-like measures. More precisely, we investigate functions within the class of convex divergences.

Divergences are often used as discrete measures in signal processing problems [7, 8]. Known examples of these functions are Kullback-Leibler (KL) [9], Jeffreys-Kullback (JK) [10], Hellinger (Hel) [11], Chi square [12] and I_α divergences. They serve as dissimilarity functions in many information theoretic models (e.g. in source and channel coding [13, 14]), data recovery tasks (e.g. image restoration [15, 16] and reconstruction [17]), machine learning (pattern recognition [18] and clustering [19]),... Note that they were used as local regularization functions [20, 21] for solving inverse problems. However, to the best of our knowledge, they have not yet been investigated for building non-local smoothness measures.

In this work, we propose a proximal optimization method for NLTV-like regularization involving divergences. We first address the problem of computing the associated proximity operators. This contribution enlarges the list of functions the proximity operator of which is given either by a closed form expression or is easily computable. Next, we develop an efficient primal-dual algorithm [22] for the restoration problem under consideration. Finally, we show experimentally the influence of various NLTV-like regularization sparsity measures on the performance of the resulting restoration method.

The remaining of the paper is organized as follows: In Section 2, we present the general form of the addressed optimization problem and introduce the notation used in this work. In Section 3, we study the proximity operators of a number of divergences. The considered restoration problem and our algorithm are then described in Section 4. Simulations are performed in Section 5, showing the good performance of the proposed approach. Finally, Section 6 concludes the paper.

2. OPTIMIZATION PROBLEM

2.1. Problem statement

In the context of inverse problems, we aim at solving problems of the form:

Problem 2.1

$$\underset{x\in\mathbb{R}^N}{\text{Minimize}}\;\; D(Ax,Bx)+\sum_{s=1}^{S} R_s(T_s x), \tag{3}$$

where $D \in \Gamma_0(\mathbb{R}^P \times \mathbb{R}^P)$, $A \in \mathbb{R}^{P\times N}$ and $B \in \mathbb{R}^{P\times N}$, and, for every $s \in \{1,\ldots,S\}$, R_s is a function in $\Gamma_0(\mathbb{R}^{K_s})$ and $T_s \in \mathbb{R}^{K_s\times N}$.

Note that functions $(R_s)_{1\le s\le S}$ may take $+\infty$ value, e.g. finite values can be assigned only to nonnegative-valued arguments. Consequently, convex constrained optimization problems can be viewed as special cases of Problem 2.1. In such cases, some of the functions $(R_s)_{1\le s\le S}$ are the indicator functions of some nonempty closed convex sets. Recall that the indicator function ι_C of a nonempty closed convex subset C of a Hilbert space $\mathcal{H}$ is defined as

$$(\forall x\in\mathcal{H})\qquad \iota_C(x)=\begin{cases}0 & \text{if } x\in C\\ +\infty & \text{otherwise.}\end{cases} \tag{4}$$

2.2. Notation and basic relations

In this paper, we deal with additive information measures of the form:

$$\big(\forall p=(p^{(i)})_{1\le i\le P}\in\mathbb{R}^P\big)\big(\forall q=(q^{(i)})_{1\le i\le P}\in\mathbb{R}^P\big)$$
$$D(p,q)=\sum_{i=1}^{P}\Phi(p^{(i)},q^{(i)}) \tag{5}$$

where Φ is defined as follows

$$\big(\forall(\upsilon,\xi)\in\mathbb{R}^2\big)$$
$$\Phi(\upsilon,\xi)=\begin{cases}\xi\,\varphi\Big(\dfrac{\upsilon}{\xi}\Big) & \text{if } \upsilon\in[0,+\infty[\text{ and } \xi\in\,]0,+\infty[\\ \upsilon\displaystyle\lim_{\zeta\to+\infty}\dfrac{\varphi(\zeta)}{\zeta} & \text{if } \upsilon\in\,]0,+\infty[\text{ and } \xi=0\\ 0 & \text{if } \upsilon=\xi=0\\ +\infty & \text{otherwise}\end{cases} \tag{6}$$

and $\varphi\in\Gamma_0(\mathbb{R})$, $\varphi\colon\mathbb{R}\to[0,+\infty]$ is twice differentiable on $]0,+\infty[$. Thus, $\Phi\in\Gamma_0(\mathbb{R}^2)$ is the perspective function [23, Chapter 3] of φ on $[0,+\infty[\,\times\,]0,+\infty[$. If φ is strictly convex and $\varphi(1)=\varphi'(1)=0$, then D belongs to the celebrated class of φ-divergences [24], i.e.

$$\big(\forall(p,q)\in[0,+\infty[^P\times[0,+\infty[^P\big)$$
$$\begin{cases}D(p,q)\ge 0\\ D(p,q)=0 \quad\Leftrightarrow\quad p=q\end{cases} \tag{7}$$

Examples of φ-divergences are listed in Table 1.

Divergence	φ
Kullback-Leibler	$\varphi(\zeta)=\zeta\ln\zeta-\zeta+1$
Jeffereys-Kullback	$\varphi(\zeta)=(\zeta-1)\ln\zeta$
Hellinger	$\varphi(\zeta)=\zeta+1-2\sqrt{\zeta}$
Chi square	$\varphi(\zeta)=(\zeta-1)^2$
I_α, $\alpha\in]0,1[$	$\varphi(\zeta)=1-\alpha+\alpha\zeta-\zeta^\alpha$

Table 1. Examples of φ-divergences

3. PROXIMITY OPERATORS OF DIVERGENCES

Definition Let $\mathcal{H}$ be a real Hilbert space endowed with the norm $\|\cdot\|$. Let $f\in\Gamma_0(\mathcal{H})$. For every $\overline{x}\in\mathcal{H}$, there exists a unique minimizer of the function $f+\frac{1}{2}\|\cdot-\overline{x}\|^2$. This minimizer is called the proximity operator of f at $\overline{x}$ and it is denoted by $\mathrm{prox}_f\overline{x}$ [25]. In other words,

$$\mathrm{prox}_f\colon\mathcal{H}\to\mathcal{H}\colon\overline{x}\mapsto\underset{x\in\mathcal{H}}{\operatorname{argmin}}\;\; f(x)+\frac{1}{2}\|x-\overline{x}\|^2. \tag{8}$$

The proximity operator has played a key role in recent developments in convex optimization, since it provides a natural extension of the notion of projection. Indeed, if C is a nonempty closed convex subset of $\mathcal{H}$, then prox_{ι_C} reduces to the projection P_C onto C. As projection operators, proximity operators are firmly nonexpansive [26], which is a fundamental property that guarantees the convergence of fixed point algorithms grounded on their use.

We are mainly concerned with the determination of the proximity of D (in this case, $\mathcal{H} = \mathbb{R}^P \times \mathbb{R}^P$). The next result shows that the problem reduces to the determination of the proximity operator of a real function of two variables.

Proximity operator One of the divergence properties is that these functions are separable (as defined in (5)). Hence, the proximity operator of D, calculated at points $\bar{p} = (\bar{p}^{(i)})_{1\le i\le P} \in \mathbb{R}^P$ and $\bar{q} = (\bar{q}^{(i)})_{1\le i\le P} \in \mathbb{R}^P$, reads

$$\mathrm{prox}_D(\bar{p},\bar{q}) = \Big(\mathrm{prox}_\Phi(\bar{p}^{(i)},\bar{q}^{(i)})\Big)_{1\le i\le P}. \tag{9}$$

Thus, in the following, we concentrate on the computation of the proximity operator of a scaled version of the involved function $\Phi \in \Gamma_0(\mathbb{R}^2)$.

Let Θ denote a primitive of the function $\zeta \mapsto \zeta\varphi'(\zeta^{-1})$ on $]0,+\infty[$ and let

$$\vartheta_-\colon\ ]0,+\infty[\ \to \mathbb{R}\colon \zeta \mapsto \varphi'(\zeta^{-1}) \tag{10}$$

$$\vartheta_+\colon\ ]0,+\infty[\ \to \mathbb{R}\colon \zeta \mapsto \varphi(\zeta^{-1}) - \zeta^{-1}\varphi'(\zeta^{-1}). \tag{11}$$

A first technical result is as follows:

Lemma 3.1 *Let $\gamma \in\ ]0,+\infty[$, let $(\overline{\upsilon},\overline{\xi}) \in \mathbb{R}^2$, and define*

$$\chi_- = \inf\big\{\zeta \in\ ]0,+\infty[\ \big|\ \vartheta_-(\zeta) < \gamma^{-1}\overline{\upsilon}\big\} \tag{12}$$

$$\chi_+ = \sup\big\{\zeta \in\ ]0,+\infty[\ \big|\ \vartheta_+(\zeta) < \gamma^{-1}\overline{\xi}\big\} \tag{13}$$

(with the usual convention $\inf\varnothing = +\infty$ and $\sup\varnothing = -\infty$). If $\chi_- \neq +\infty$, the function

$$\psi\colon\ ]0,+\infty[\ \to \mathbb{R}\colon \zeta \mapsto \zeta\varphi(\zeta^{-1}) - \Theta(\zeta) + \frac{\gamma^{-1}\overline{\upsilon}}{2}\zeta^2 - \gamma^{-1}\overline{\xi}\zeta \tag{14}$$

is strictly convex on $]\chi_-,+\infty[$. In addition, if

1. $\chi_- \neq +\infty$ *and* $\chi_+ \neq -\infty$
2. $\lim_{\substack{\zeta\to\chi_-\\ \zeta>\chi_-}} \psi'(\zeta) < 0$
3. $\lim_{\zeta\to\chi_+} \psi'(\zeta) > 0$

then ψ admits a unique minimizer $\widehat{\zeta}$ on $]\chi_-,+\infty[$, and $\widehat{\zeta} < \chi_+$.

Using Lemma 3.1, we obtain the following characterization of the proximity operator of any scaled version of Φ:

Proposition 3.2 *Let $\gamma \in\ ]0,+\infty[$ and $(\overline{\upsilon},\overline{\xi}) \in \mathbb{R}^2$. $\mathrm{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi}) \in\ ]0,+\infty[^2$ if and only if Conditions 1-3 in Lemma 3.1 are satisfied. When these conditions hold*

$$\mathrm{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi}) = \big(\overline{\upsilon} - \gamma\,\vartheta_-(\widehat{\zeta}), \overline{\xi} - \gamma\,\vartheta_+(\widehat{\zeta})\big) \tag{15}$$

where $\widehat{\zeta} < \chi_+$ is the unique minimizer of ψ on $]\chi_-,+\infty[$.

The existence of $\widehat{\zeta}$ being guaranteed, it can be computed by standard one-dimensional search techniques, which are implementable in parallel. Due to the limited space, the reader is referred to [27] for further details.

4. APPLICATION TO IMAGE RESTORATION

A common problem in image restoration is to recover an original image $\overline{x} \in \mathbb{R}^N$ from an observation vector $z \in \mathbb{R}^Q$, where

$$z = H\overline{x} + w, \tag{16}$$

$H \in \mathbb{R}^{Q\times N}$ is a linear operator modeling some blur and $w \in \mathbb{R}^Q$ is a realization of an additive zero-mean white Gaussian noise. In some instances, $\overline{x}$ can be estimated from z be employing the least squares criterion $x \mapsto \frac{1}{2}\|Hx - z\|^2$. However, as inverse problems are usually ill-posed, one needs some prior information about $\overline{x}$. This additional information is reflected in the resulting optimization problem by a regularization term, e.g. a NLTV-like term, which serves to control the smoothness of the estimate. Therefore, the restoration can be achieved by solving the following convex optimization problem:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \frac{1}{2\lambda}\|Hx - z\|^2 + D(Ax,Bx) + \iota_C(x) \tag{17}$$

where $\lambda \in]0,+\infty[$ is the regularization constant, and the operators A and B are chosen as explained in the introduction. The last term ι_C constrains x to belong to the convex set $C = [0,255]^N$.

Alternatively, in this paper, we seek to describe the problem within a set theoretic framework. In several works [28], it was observed that an upper bound on the data fidelity term allows us to efficiently restrict the solution to vectors x such that:

$$Hx \in C' = \big\{u \in \mathbb{R}^Q \mid \|u - z\|^2 \le \delta Q\sigma^2\big\} \tag{18}$$

where δ is a positive constant (usually close to 1) and σ^2 is the noise variance. This leads to the following variant of Problem (17):

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ D(Ax,Bx) + \iota_{C'}(Hx) + \iota_C(x). \tag{19}$$

Note that, under technical assumptions, Problem (19) is equivalent to Problem (17). However the constrained formulation given above is usually considered to be more practical as the solution is less sensitive to the choice of δ than λ [4].

The above problem can be solved using proximal optimization algorithms. Such methods require to compute the proximity operator of the divergence D, which has already been discussed in Section 3, the projection onto the ℓ_2 ball, and the projection onto the hypercube $[0,255]^N$. These two projections are quite standard. A possible algorithm for solving Problem (19) is thus the M+LFBF primal-dual algorithm [22]. The associated iterations are recalled in Algorithm 1, where at each iteration k, $\gamma^{[k]}$ is a step-size and $e^{[k]} \in (\mathbb{R}^P)^2$ corresponds to a possible error in the computation of the proximity operators of the divergence term.

Algorithm 1: M+LFBF

Initialization

$$v_1^{[0]} \in \mathbb{R}^P, v_2^{[0]} \in \mathbb{R}^P, v_3^{[0]} \in \mathbb{R}^Q, x^{[0]} \in \mathbb{R}^N$$
$$\beta = \left(\|A\|^2 + \|B\|^2 + \|H\|^2\right)^{1/2},\ \varepsilon \in]0, 1/(\beta+1)[$$

For $k = 0, 1, \ldots$

$$\begin{aligned}
&\gamma^{[k]} \in [\varepsilon, (1-\varepsilon)/\beta] \\
&y_1^{[k]} = x^{[k]} - \gamma^{[k]}(A^\top v_1^{[k]} + B^\top v_2^{[k]} + H^\top v_3^{[k]}) \\
&p_1^{[k]} = \mathsf{P}_C(y_1^{[k]}) \\
&(y_{2,0}^{[k]}, y_{2,1}^{[k]}) = (v_1^{[k]}, v_2^{[k]}) + \gamma^{[k]}(Ax^{[k]}, Bx^{[k]}) \\
&(p_{2,0}^{[k]}, p_{2,1}^{[k]}) = (y_{2,0}^{[k]}, y_{2,1}^{[k]}) \\
&\qquad\qquad - \gamma^{[k]} \mathrm{prox}_{\frac{D}{\gamma^{[k]}}}\left(\frac{y_{2,0}^{[k]}}{\gamma^{[k]}}, \frac{y_{2,1}^{[k]}}{\gamma^{[k]}}\right) + e^{[k]} \\
&(q_{2,0}^{[k]}, q_{2,1}^{[k]}) = (p_{2,0}^{[k]}, p_{2,1}^{[k]}) + \gamma^{[k]}(Ap_1^{[k]}, Bp_1^{[k]}) \\
&(v_1^{[k+1]}, v_2^{[k+1]}) = (v_1^{[k]}, v_2^{[k]}) - (y_{2,0}^{[k]}, y_{2,1}^{[k]}) + (q_{2,0}^{[k]}, q_{2,1}^{[k]}) \\
&y_3^{[k]} = v_3^{[k]} + \gamma^{[k]} H x_3^{[k]} \\
&p_3^{[k]} = y_3^{[k]} - \gamma^{[k]} z - \gamma^{[k]} \mathsf{P}_{C'}\left(\frac{y_3^{[k]}}{\gamma^{[k]}} - z\right) \\
&q_3^{[k]} = p_3^{[k]} + \gamma^{[k]} H p_1^{[k]} \\
&v_3^{[k+1]} = v_3^{[k]} - y_3^{[k]} + q_3^{[k]} \\
&q_1^{[k]} = p_1^{[k]} - \gamma^{[k]}(A^\top p_{2,0}^{[k]} + B^\top p_{2,1}^{[k]} + H^\top p_3^{[k]}) \\
&x^{[k+1]} = x^{[k]} - y_1^{[k]} + q_1^{[k]}.
\end{aligned}$$

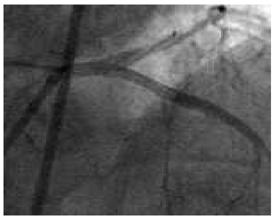
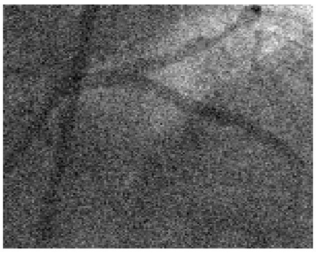


Fig. 1. Original image ”Aniso“.

5. RESULTS

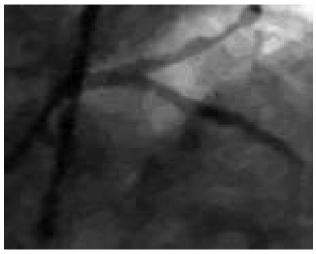
In this section we present the performance of the proposed D-NLTV regularization method on restoration experiments. In particular, the following choices for D are considered: ℓ_2 norm (as usually considered in the literature), Kullback-Leibler divergence, Hellinger divergence, and Chi square divergence. In our experiments, $\overline{x}$ corresponds to a 150×150 angiographic image from the public domain (www.pcronline.com) (Fig. 1). The observed image is generated by degrading the original image with a convolution operator H, which is equal to identity for the denoising problem and corresponds to a truncated Gaussian point spread function with standard deviation 1.6 and kernel size 3×3 for the deconvolution problem. The noise variance is equal to 400 and 64 for the image denoising and restoration problems, respectively. The linear operators A and B associated with NLTV are computed from the TV image result obtained using the code in [29] with $P_n \equiv 10$. The balance between the smoothness of the estimate and the data fidelity is controlled by the parameter δ tuned so as to maximize the Signal-to-Noise Ratio (SNR). The quality of the results, presented in Figs 2 and 3, is evaluated in terms of the SNR, the Mean Absolute Error (MAE), and the Structural Similarity index (SSIM) [30]. One can observe that the results obtained with D-NLTV outperform the standard ℓ_2-NLTV, for the different divergences. Some isolated noisy points in the result corresponding to ℓ_2-NLTV prior (Fig. 2c) are visible. These isolated noisy pixels have been successfully removed when using divergence based criteria.



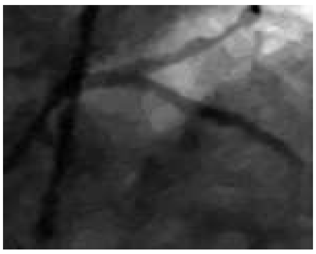
(a) Noisy (13.73, 16.04, 0.250)

(b) TV (24.55,4.36,0.830)

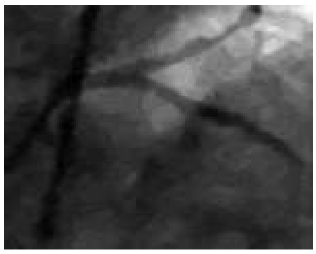
(c) ℓ_2-NLTV (24.88,4.20,0.830)



(d) KL-NLTV (25.11,4.12,0.843)



(e) Hel-NLTV (25.09, 4.13, 0.843)



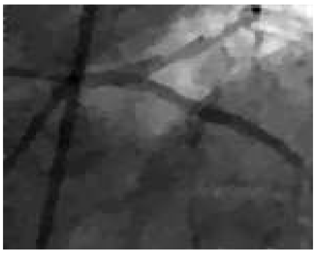
(f) Chi-NLTV (25.10,4.12,0.843)

Fig. 2. Denoising problem results (SNR, MAE, SSIM).

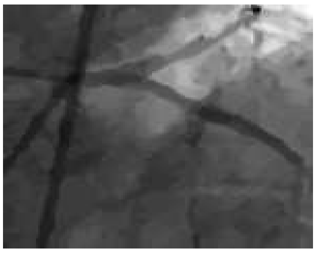
6. CONCLUSION

In this paper, we have proposed a convex optimization framework involving various information divergences. We have provided the expression of the proximity operators of these divergences, which facilitates the use of these measures in inverse problems encountered in signal and image processing. As a side result, we presented the performance obtained by employing these divergences as non-local regularity measures for restoration problems, in the presence of Gaussian noise.

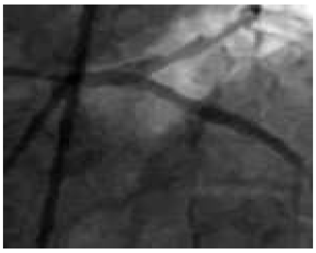
(a) Degraded (19.88, 7.25, 0.605)



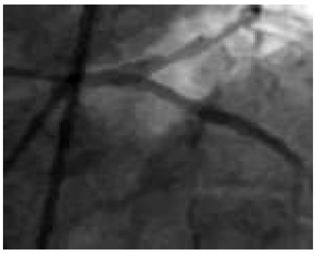
(b) TV (27.26,3.17,0.885)



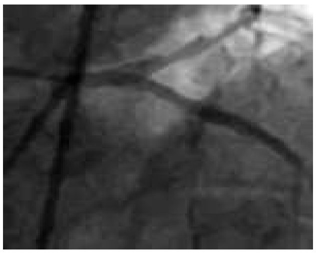
(c) ℓ_2-NLTV (27.63,3.06,0.888)



(d) KL-NLTV (27.77,3.01,0.894)



(e) Hel-NLTV (27.75, 2.99,0.894)



(f) Chi-NLTV (27.76, 3.01, 0.894)

Fig. 3. Restoration problem results (SNR, MAE, SSIM).

7. REFERENCES

[1] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.

[2] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, San Francisco, CA, USA, 2010, pp. 2464–2471.

[3] G. Peyré, "A review of adaptive image representations," *IEEE J. Selected Topics Signal Process.*, vol. 5, no. 5, pp. 896–911, 2011.

[4] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems: Part I," 2012, http://arxiv.org/abs/1210.5844.

[5] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot, "Dual constrained TV-based regularization on graphs," *SIAM J. on Imaging Sciences*, pp. 1–27, Oct. 2013, http://hal.archives-ouvertes.fr/hal-00743968/.

[6] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, San Diego, CA, USA, 2005, pp. 60–65.

[7] W. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am. A*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[8] T. Teuber, G. Steidl, and R. H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Preprint University of Kaiserslautern*, 2012.

[9] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[10] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.

[11] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.," *Journal für die reine und angewandte Mathematik*, vol. 136, pp. 210–271, 1909.

[12] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonable be supposed to have arisen from random sampling," *Philosophical Magazine*, vol. 50, pp. 157–175, 1900.

[13] R.E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[14] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[15] F.-X. Dupé, J.M. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310 –321, Feb. 2009.

[16] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel proXimal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.

[17] J.A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1439 –1450, Oct. 1995.

[18] W.-L. Hung and M.-S. Yang, "On the J-divergence of intuitionistic fuzzy sets with its application to pattern recognition," *Inf. Sci.*, vol. 178, no. 6, pp. 1641–1650, Mar. 2008.

[19] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, Dec. 2005.

[20] S. Joshi and M.I. Miller, "Maximum α posteriori estimation with Good's roughness for three-dimensional optical-sectioning microscopy," *J. Opt. Soc. Am. A*, vol. 10, no. 5, pp. 1078–1085, May 1993.

[21] J. A. O'Sullivan, "Roughness penalties on finite domains," *IEEE Trans. Image Process.*, vol. 4, no. 9, pp. 1258–1268, Sep. 1995.

[22] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, vol. 20, no. 2, pp. 307–330, Jun. 2012.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[24] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[25] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.

[26] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM J. on Optimization*, vol. 18, no. 4, pp. 1351–1376, Nov. 2007.

[27] J.-C. Pesquet and M. El-Gheche, "Proximity operators of some discrete information divergences," Tech. Rep., Paris-Est University, 2012.

[28] P.L. Combettes and H.J. Trussell, "The use of noise properties in set theoretic estimation," *IEEE Trans. Signal Process.*, vol. 39, no. 7, pp. 1630–1641, Jul. 1991.

[29] X. Bresson, "A short note for nonlocal TV minimization," Jun. 2009, `http://www.cs.cityu.edu.hk/~xbresson/codes.html`.

[30] W. S. Malpica and A. C. Bovik, "Range image quality assessment by structural similarity," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Washington, DC, USA, 19-24 Apr. 2009, pp. 1149–1152.